



Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

Elsevier hereby grants permission to make all its COVID-19-related research that is available on the COVID-19 resource centre - including this research content - immediately available in PubMed Central and other publicly funded repositories, such as the WHO COVID database with rights for unrestricted research re-use and analyses in any form or by any means with acknowledgement of the original source. These permissions are granted for free by Elsevier for as long as the COVID-19 resource centre remains active.

Virus Evolution and Genetics

OUTLINE

Virus Evolution	81	<i>Forward and Reverse Genetics</i>	84
Virus Genetics	84	<i>Reverse Genetics of Positive-Strand RNA Viruses</i>	85
<i>Recombination</i>	84	<i>Reverse Genetics of Negative-Strand RNA Viruses</i>	85
<i>Genotypes and Phenotypes</i>	84	<i>Reverse Genetics of Retroviruses</i>	85
<i>Nonessential Genes</i>	84	<i>Reverse Genetics of DNA Viruses</i>	85

After reading this chapter, you should be able to discuss the following:

- Did a single common ancestor give rise to all viruses?
- What are the two discrete processes that drive virus evolution?
- How is the mutation rate of a virus described?
- How is the evolution rate of a virus described?
- What are “nonessential genes” and are they always nonessential?
- What is “reverse genetics” and how is it used to probe the functions of virus genes?

In the first part of this chapter we discuss (in very general terms) when and how viruses originated and then consider the ongoing process of virus evolution. In the second part of this chapter we discuss the molecular genetics of some major groups of viruses.

VIRUS EVOLUTION

There is great diversity among viruses; they infect organisms from all Kingdoms of life and across all ecosystems. Animals, plants, single-celled eukaryotes, and prokaryotes are infected by a variety of viruses, both DNA and RNA. And some closely related viruses infect animals, plants, and bacteria. These facts suggest a very ancient origin for viruses.

There are three commonly proposed mechanisms for the origins of viruses:

- *Viruses descended from primitive precellular life forms.* This theory posits that viruses originated and evolved along with the primitive self-replicating molecules that were destined to become cells; the agents we recognize as viruses today were originally self-replicating molecules in the precellular world. If this is correct, it follows that cellular life forms were impacted by “viruses” from their earliest beginnings.
- *Viruses are “escaped” cellular genetic elements.* This theory posits that viruses evolved *after* cells. Their origins were cell-associated genetic elements that acquired protein coats, allowing for more efficient cell to cell transfer. The DNA genomes of some viruses do resemble plasmids, and retroviruses are related to the large group of nonviral retro-elements populating cellular genomes.
- *Retrograde evolution.* The theory of retrograde evolution states that viruses were once complex intracellular parasites that lost the ability for all independent metabolism; they retained only those genes required to manipulate the host cell and produce progeny virions.

It may be that each theory applies to a different viral lineage. For example, it is commonly thought that the first replicating molecules were RNA. Thus RNA viruses could be descended from these molecules.

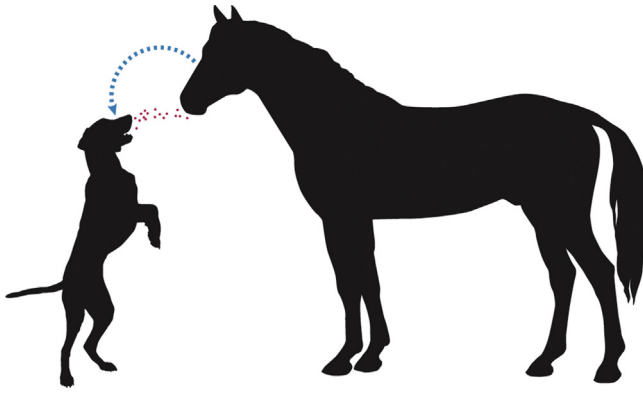


FIGURE 8.1 Virus moves to a new species: In 2004 equine influenza virus jumps to dogs.

As cellular life evolved, replicating RNA molecules would need to parasitize cells to access pools of biomolecules. The “escaped cellular element” theory could explain the origins of retroviruses, and the largest DNA viruses may have descended from more complex parasites. In contrast, some evolutionary biologists see evidence that large DNA viruses are acquiring genes from their hosts.

Regardless of the origins of viruses, there is good evidence that they have shaped the evolution of single-celled and multicellular organisms for billions of years. Animal genomes are littered with pieces of RNA and DNA genomes that have very ancient origins and there are good examples of animals coopting viral genes. For example, genes necessary for the development of the mammalian placenta were obtained from retroviruses and some animal cells use captured viral gene fragments to defend against further infection!

While some virologists ponder the ancient origins of viruses, others examine the mechanisms that drive ongoing virus evolution. Examples of ongoing virus evolution include:

- Cross-species jumps that allow viruses to find new hosts. Influenza viruses (Chapter 23: Family *Orthomyxoviridae*) provide many examples of this type: Bird or swine influenza jumped into humans at the turn of the 20th century and horse influenza virus moved into dogs in 2004 (Fig. 8.1). Among the DNA viruses, a dog-adapted version of a feline parvovirus emerged in the 1970s and rapidly spread worldwide through the domestic dog population. Ongoing transmission in the new host is a measure of successful adaptation.
- Decreased virulence in a new host. Viruses that are well adapted to their hosts often cause little or no disease. When a virus jumps to new host, the virus

may cause severe disease, but coevolution of virus and host often restores more balance. An instructive example is provided by studies of rabbit myxomavirus (Chapter 35: Family *Poxviridae*).

- Emergence of drug resistant viruses. As antiviral drugs have been developed, drug resistant viruses have emerged. The mutation and replication rates of human immunodeficiency virus (HIV) are so high that successful long-term treatment requires use of drug cocktails, containing three or four different active compounds (to minimize development of drug resistance).
- Immune escape mutants. HIV and influenza viruses are adept at escaping from host immune pressures. Point mutations in certain regions of influenza virus surface proteins [hemagglutinin (HA) and neuraminidase (NA)] are well tolerated by the virus but can thwart established immune responses. This is one reason that influenza viruses are closely monitored and vaccines are updated yearly (Chapter 23: Family *Orthomyxoviridae*).

Virus evolution is the outcome of two independent events (Fig. 8.2): The first event is mutation of the viral genome. For RNA viruses, this is often a frequent event during genome replication, as most RNA-dependent RNA polymerases have no proof reading functions (this is why some RNA viruses exist as a quasispecies as described in Chapter 10: Introduction to RNA Viruses. DNA damage may play a role in mutation of some DNA virus genomes. Virus genomes can also mutate as the result of recombination or even by capturing genes from the host or from other viruses!

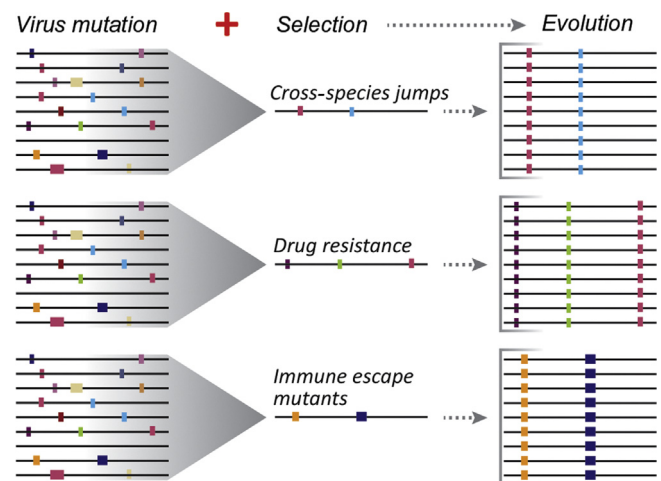


FIGURE 8.2 Evolution is the result of two independent processes: Mutation and selection.

There are several ways to express mutation rates. A common method is to express them as misincorporations per nucleotide synthesized. Another way is to express them as mutations/nucleotide/cell infection as this accounts for different modes of viral genome replication. Accepted estimates range from 10^{-6} to 10^{-4} mutations/nucleotide/cell infection for RNA viruses and range from 10^{-8} to 10^{-6} mutations/nucleotide/cell infection for DNA viruses. For some RNA viruses, these numbers translate into one or two mutations per genome replicated. Finally, viral mutation rates can be difficult to measure, as lethal mutations are quickly eliminated from the population.

The second event that drives virus evolution is *selection* of mutants that are more fit for a particular environment. For example, beneficial mutations may allow viruses to expand their host-range, evade immune responses, or tolerate antiviral drugs. Thus virus evolution requires initial mutations followed by selective pressures. A measurement of virus evolution that includes the parameter of natural selection is nucleotide substitutions per nucleotide site, *per year*. This measure reflects mutations that become *fixed* in the virus population. If a virus is highly adapted to a particular host and environment, this number may be very low. In contrast, the rate may be much higher if a virus is in the process of adapting to a new host or environment. Different lineages of the same virus may evolve at different

rates depending on the selective pressures they encounter.

One common question about viruses is their “age.” Is human immunodeficiency virus (HIV) a few decades old, a few hundred years old or many millions of years old? While we know that the virus we call HIV-1 *emerged* in the human population several decades ago, it is very closely related to viruses of nonhuman primates that are millions of years old. There are different methods for calculating rates of virus evolution (that will not be described here) and these can provide very different answers about the age of a virus. A relatively recent finding that impacts our view of virus age is that animal genomes harbor pieces of viral genomes (a variety of DNA and RNA viruses in addition to retroviruses) that are tens or hundreds of millions of years old. We can time the insertion events by looking at closely related species of animals. If nearly identical virus insertions are present in two different species, it is assumed that the viral gene insertion occurred prior to the split of those two species. Very surprising is the ease with which we recognize these so-called “endogenized” or “endogenous” viruses as *clearly related* to viruses widely circulating today. Research is finding that in some cases, the endogenized virus genes are expressed, and may serve to protect cells from infection with currently circulating viruses. In addition there are well characterized examples of co-evolution among organisms and viruses (Box 8.1).

BOX 8.1

COEVOLUTION OF VIRUS AND HOST

Animals have evolved a variety of strategies to control/limit/inhibit virus replication. In response, viruses have evolved mechanisms to “fight back.” The ability to readily sequence both animal genomes and viruses is providing a detailed view into specific instances of evolution and coevolution of viruses and their hosts. The story we consider here is that of tetherin and its virally encoded antagonists.

Tetherin is a potent antiviral protein that inhibits a variety of enveloped viruses, including retroviruses. Tetherin expression is induced by interferon in response to viral infection (Chapter 6: Immunity and Resistance to Viruses). As the name implies, tetherin inhibits release of enveloped viruses from the cell surface (it “tethers” them to the cell). Tetherin appears to physically crosslink virions to the plasma membrane.

How then do viruses antagonize tetherin? The retroviruses simian immunodeficiency virus SIVcpz (from chimpanzees) and SIVgor (from gorillas) antagonize their hosts’ tetherin with a protein called NEF [Chapter 37: Replication and Pathogenesis of Human Immunodeficiency Virus (HIV)]. SIV NEF binds to a specific amino acid in monkey tetherin to counteract its activity. However, human tetherin has a deletion where NEF would bind. Therefore SIVs replicate poorly in humans because human tetherin *is not* antagonized by SIV NEF.

So how then is HIV able to replicate to high levels in humans? HIV evolved a completely different tetherin antagonist! HIV encodes a protein called viral protein U (VPU) that counteracts tetherin by interacting at an entirely different site. HIV still encodes a NEF protein (it has a number of different functions that help HIV replicate) but it uses VPU to antagonize tetherin.

VIRUS GENETICS

Recombination

The principles and terminology used to describe the genetics of cellular organisms are applicable to viral genetics as well. In the sections above we used the terms mutation, selection, and recombination with the assumption they would be understood by students of biology. However recombination of virus genomes merits particular mention as it includes some unique mechanisms. Homologous recombination between DNA virus genomes occurs by the same molecular mechanisms described for cellular DNA. However RNA viruses also undergo genome “recombination” by a different molecular mechanism. RNA virus recombination is achieved through a process called “copy choice.” Copy choice involves template switching during genome replication. The RNA replicase complex physically dissociates from one genome and reassociates with another. Regions of genome complementarity probably serve to position the replicase complex on the new template. The frequency of copy choice recombination is dependent upon the mechanisms used by a particular virus family for genome replication and transcription. As described in Chapter 17, Family *Coronaviridae*, coronaviruses synthesize their mRNA by a process of discontinuous transcription, whereby the transcription complex jumps from one position on the genome to another. It is thought that this mechanism results in frequent template switching during genome replication as well, resulting in frequent recombination events.

Another mechanism of “recombination” comes into play for segmented RNA viruses (i.e., orthomyxoviruses, reoviruses, and bunyaviruses). The orthomyxoviruses (influenza viruses) have segmented genomes (eight segments for influenza A viruses) and upon infection of a cell with two different viruses an assortment of progeny can be released (Fig. 23.8). This process is called *reassortment* as it involves shuffling of complete genome segments. Particular genome segments can also undergo copy choice recombination but reassortment is a more common event.

Genotypes and Phenotypes

The terms genotype and phenotype apply to viruses. The viral genotype is simply the sequence of its viral genome. But what kind of virus phenotypes can be measured? Quite an array as it turns out!

Virus phenotypes include: the ability to productively infect different cell types (host range); the ability to replicate at elevated temperatures; the ability to form different types of plaques (large, small, clear, or

cloudy); the ability to transform cells; to replicate at an increased or decreased rate; to cause disease; to be neutralized by a particular antibody; to replicate in the presence of various drugs; to display increased or decreased mutation and/or recombination rates.

Nonessential Genes

Examining the genetics of viruses in *cultured cells* led to a surprising finding: Not all viral genes appeared to be essential for replication. In fact, some of these so-called nonessential genes were initially identified because they were lost during passage in cultured cells, and the mutated viruses outcompeted their wild-type parents.

It was counterintuitive to some virologists to think that viruses, with their very ancient origins, maintained nonessential genes for millions of years. They hypothesized that while some genes are not essential in cultured cells, they must be advantageous under more exacting conditions, such as replication and spread in a natural population of susceptible hosts. And indeed, this appears to be the case. Over and over again, genes identified as nonessential in cultured cells are indeed required to maintain the virus in an animal population. For example, the herpesvirus thymidine kinase gene is nonessential for virus replication in cell culture but is essential for productive infection of natural hosts.

The loss of nonessential genes in cultured cells is virus evolution in action. Given an environment without physical barriers to infection, antiviral responses, or limited nutrients viruses discard “useless” genes. The benefit is a smaller genome that it can be replicated at a faster rate, using fewer resources. Thus viruses have a set of essential genes, required under all conditions (for example, capsid proteins, polymerases, replication origin binding proteins) and other genes that may only be required in a specific ecological niche. A gene may not be *absolutely* required in a particular niche, but its presence may confer a strong advantage. Thus not only is host environment important, so too are the presence or absence of viral competitors!

Forward and Reverse Genetics

The classical study of virus genetics (also called “forward” genetics) involved identifying a novel phenotype and then identifying the associated mutations. To accomplish this, virus stocks were subject to mutagenesis and large numbers of mutants were analyzed individually. A variety of methods were used to sort out and classify mutants. The approach worked well for bacteriophages (bacterial viruses) and it was also

applied to animal viruses when methods were developed to grow and plaque purify them using cultured cells. However, it can be challenging to design strategies for selecting a desired phenotype, and brute screening can take years.

Today, studies of viral genetics depend largely on a general methodology called *reverse genetics*. The term reverse genetics is used to describe activities that involve *making specific mutations* in a viral genome and *then observing* the phenotype. Mutations can range from a single nucleotide substitution to partial or complete gene deletions. Sometimes the order or arrangement of genes is altered. The goal of these changes is to link gene function to altered phenotypes, thereby generating models that describe gene function. Most significant human/animal viral pathogens can be studied using reverse genetic approaches. Perhaps one of the most impressive (and controversial) feats in reverse genetics was the resurrection of the deadly 1918 pandemic strain of influenza virus, using genome sequences derived from decades-old, formalin-fixed lung autopsy materials and frozen tissues from victims buried in permafrost.

Reverse genetic analysis involves both *design* of a mutation and the ability to *incorporate* the mutation into a virus. Generating gene mutations uses a variety of methods available through modern molecular biology. (If anyone doubts the utility of “basic research” consider the powerful toolkit it has provided for the study of human and animal pathogens, genetic disorders, and cancer.) In the study of viruses, principles governing both the design of mutations and their incorporation into viruses benefit from having detailed knowledge of viral genome structure and replication strategy. As will be seen in the coming chapters, these vary widely, but there are some common principles.

Reverse Genetics of Positive-Strand RNA Viruses

The *genomes* of positive sense RNA viruses are infectious in the absence of any viral proteins (Chapter 10: Introduction to RNA Viruses). The infectious “heart” of a positive-strand RNA virus is its genome. Therefore the genomes of many positive-strand RNA virus can be synthesized and introduced into a cell with the aim of producing virions. The general process is outlined in Fig. 8.3.

Reverse Genetics of Negative-Strand RNA Viruses

The *genomes* of negative sense, ambisense and double-stranded RNA viruses are not infectious in the

absence of viral proteins (Chapter 10: Introduction to RNA Viruses). In a natural infection, the genome is associated with viral proteins, key among them an RNA-dependent RNA polymerase. Therefore one must introduce, or express within a cell, both a viral genome and a subset of viral proteins before the replication cycle to produce virions can begin.

Reverse Genetics of Retroviruses

The genomes of retroviruses are mRNA (Chapter 36: Family *Retroviridae*). However they do not replicate in the manner of positive-strand RNA viruses, as their RNA genomes are not translated to initiate a productive infection. Instead, early in the infection process, retroviruses synthesize of a DNA copy of their RNA genome. Therefore, reverse genetics of retroviruses uses plasmids containing the DNA copies of retroviral genomes. The retroviral genome is expressed (transcribed) in the cell nucleus to produce the set of mRNAs needed to drive virion production.

Reverse Genetics of DNA Viruses

The genomes of many DNA viruses are infectious and they can be synthesized, manipulated, and introduced back into cells as plasmids, with the resulting production of virions. However, some DNA viruses also require a set of proteins to initiate an infection. Herpesvirus (Chapter 34: Family *Herpesviridae*) and poxvirus (Chapter 35: Family *Poxviridae*) virions contain proteins needed to initiate a productive infection cycle. Introducing mutations into these viruses depends on homologous recombination within an infected cell to exchange specific genome segments.

In this chapter we learned that:

- Two separate processes, mutation and selection are the drivers of virus evolution.
- Virus *mutation rates* are described in several ways, such as mutations per nucleotide synthesized, mutations per genome synthesized or mutations/nucleotide/cell infection. For some RNA viruses, these numbers translate into one or two mutations per genome replicated. Estimates may try to account for the length of the virus replication cycle, the number of genomes produced, or the method of genome synthesis.
- The *evolution rate* of a virus is often described as nucleotide substitutions per nucleotide site, *per year*.
- Viruses have core sets of essential genes that are always required for genome replication. Examples include capsid proteins and polymerases. Other genes, sometimes called nonessential, are dispensable, and may be lost, under certain specific

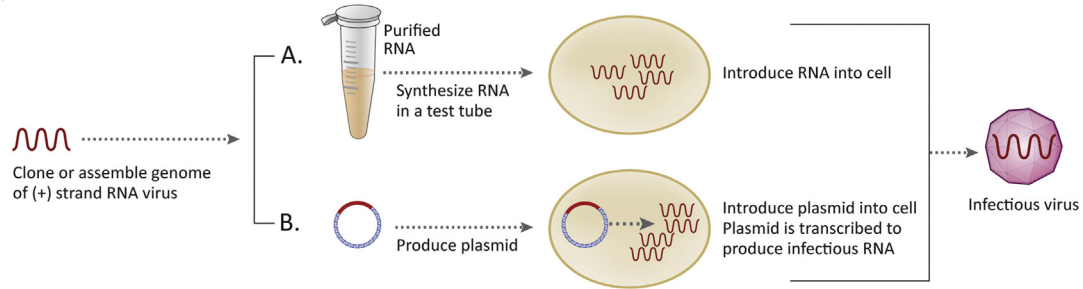
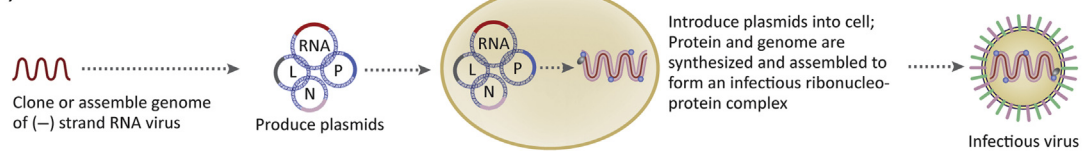
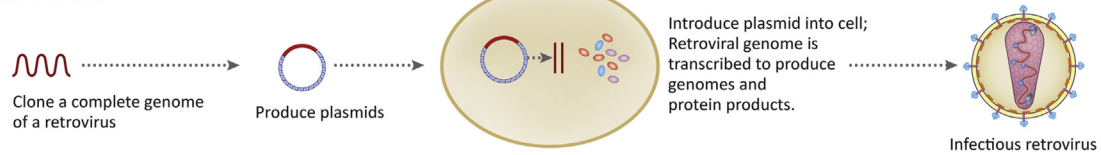
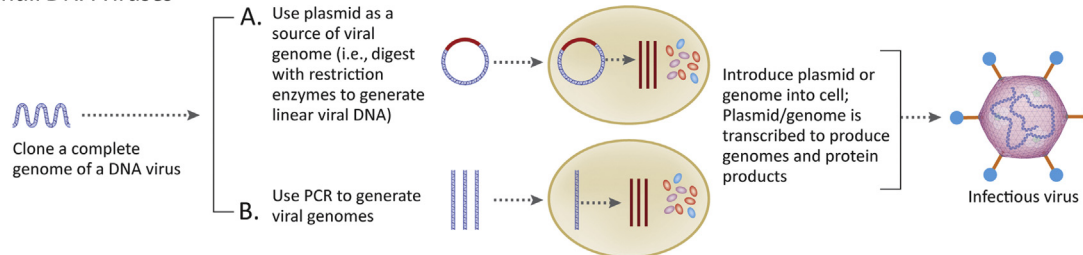
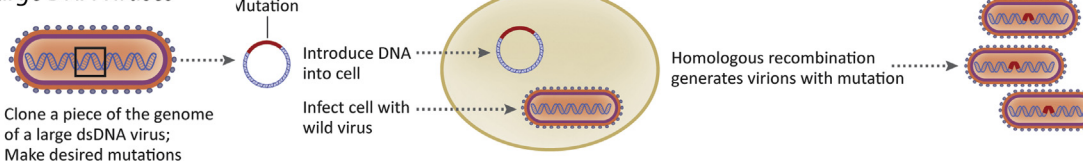
(+) Strand RNA viruses*(-) Strand RNA viruses**Retroviruses**Small DNA viruses**Large DNA viruses*

FIGURE 8.3 Reverse genetics systems for some common DNA and RNA viruses.

- some conditions (for example, during replication in cultured cells).
- The process of reverse genetics is used to determine the function of viral genes. It involves cloning a

viral genome, designing and making specific mutations, and reintroducing the genome into cells to obtain mutated virions whose phenotype can be determined.